Do We Need to Understand the World to Know It? Knowledge in a Big Data World

In both practice and academia, the incredible access to data is creating a marked disruption. The perfect storm of massive amounts of digital data, sophisticated analytical tools and cheap, scalable processing power has fostered a data-driven mindset in both corporate and academic practices. These are being embraced to varying degrees in companies around the world as well as in global academic communities. However, swinging the pendulum in the relationship between knowledge and data too far toward data can have adverse consequences. This editorial offers a cautionary note for both companies and academia.

For a number of years, in practice, the distinction between data, information, and knowledge could be stated with some level of precision. Data are raw facts and figures that can become information when massaged and placed in the right context. The value-adding activities from data to information are largely the domain of information systems. Knowledge, however, adds experience and expertise to the information and often resides in tacit form in people’s heads. So, the Knowledge Management System (KMS) popularity that started in the 1990s was intended to capture knowledge (largely tacit) and put it in a system that could benefit others in the organization (Davenport & Grover, 2001). KM processes include externalization (taking tacit knowledge and representing it in a KMS) and internalization (making this knowledge accessible to people who might need it). For example, a global consulting firm might have a team that concluded a multi-year project in Malaysia – and their experiences, successes, failures, precautions, and guidance in a KMS – could be invaluable for other teams initiating projects in that region. KM practices included creating the right incentives for knowledge to flow between people and the system, as well as embedding knowledge into products and services offered. In academic research, particularly the social sciences including Information Systems, knowledge is largely represented as an abstraction – a theory or model – that explains or predicts the real world. It is in the efficacy of these abstractions, as assessed through logic or impact, that academics gain their esteem.

The advent of “big data” in recent years, along with advanced analytics and machine learning has created a disruption both in the practice of KM as well as in the importance of abstraction. Big data, through its global reach and its sheer volume, velocity, and variety, along with computationally intensive analysis, offers opportunities for generating new insights. Mining of big data and digital streams can yield fresh perspectives for decision-makers, optimize and automate processes, and discover new ways to understand and fulfill customers’ needs. It can also offer precision of predictions to important questions in both practice and academia. However, in observing the ready embrace of big data and analytics there are also signs that some companies and researchers might fall into a trap where they see data replacing knowledge, blurring the distinctions in the trichotomy described above. The argument implicit in this, is that knowledge was largely a correspondence between our observation of the world and its interpretation in our brain, in the form of models that we used to make sense of observations. Now, with the plethora of data, we do not need to understand the world to know it. All knowledge can be extracted through the data. So, in this view, knowledge is not internalized through human assessment but externalized through data.

While many may not endorse the data is knowledge view, companies that intensively invest in data and analytical capabilities could be accepting a data culture that comes at the cost of human interpretation and judgment. Companies, where everything revolves around data science and “show
me the data” is the prevailing edict, might be overshooting the mark. There are caveats with equating big data management with knowledge management. We present three, but there could be more.

First, big data and analytics are largely about prediction; knowledge is about explanation and causality. So, the presumption is that if data can be used to predict accurately, then there is little reason to understand why the prediction works. While this could be true for machine learning algorithms that read an MRI, the vast majority of businesses deal with customers and employees, and human behaviors. Often big data sets are digitally captured (e.g. on apps, websites) for transactional purposes, and not for predicting certain outcomes. Therefore, such data sets are replete with correlates that could yield high predictive accuracy but be fallacious. Shark attacks are good predictors of ice-cream sales on the beach – which could lead to ridiculous decision for an ice-cream seller. An obvious correlate in this confounding is the temperature. However, to decipher this, a data set is needed that includes that correlate, and expertise is needed to identify the correlation-causality fallacy.

Second, big data by definition is historical … and subject to biases. Predictions are only as good as the data. So, predicting the profile of effective US Senators based on historical data on prior senators will inevitably have “male” in the profile as the dominant group in the population, reinforcing prior prejudices. Similarly, evaluations of products and services could be due to herd effects, where users provide good ratings because other ratings are positive. Understanding limitations of prior data requires human knowledge, expertise, and experience in the domain of interest.

Thirdly, the data is often subject to the “streetlight effect” where access has primacy over the questions asked. Instead of developing strategic questions, the inclination would be to mine the data to see what questions can be addressed. Or, a manager might simply use accessible data over having a systematic data capture strategy. For example, a supermarket chain manager engages in data mining to identify per-product profitability using historical sales and cost data. Then, based on the results, a decision is made to purge over a dozen product lines. The decision might be absolutely wrong, if a broader data set that examined shopping behaviors of customers who purchased these products on a single shopping trip. Such data might reveal that some products being purged could be highly profitable as high-end customers were drawn to the store due to that product and bought a number of high-margin products on the same trip (Lyytinen & Grover, 2017). Here again, understanding and domain experience would have been invaluable in asking better questions and spreading a wider data net.

In the academic research sphere, the same caveats apply. However, here, the issue is not regarding the equating of big data with KM, but more about big data approaches obviating the need for abstraction in the form of theory. For instance, in the IS field, there is a growing constituency of researchers that believe that strong analytics approaches on big data that result in accurate predictions allows us to achieve the ends directly without the need for messy abstractions or theories to muddy the water (Johnson, Gray, & Sarkar, 2019). After all, they argue, is not the accuracy of the prediction the most important practical value we can offer in our research? The counter to that are the three caveats I describe above. All of them apply to academic research – but here, there are broader issues of concern. Big data, often digital, is not collected for academic problems – it can at best capture variables that offer tactical insight to companies. So problems that emerge from big data sets are tactical and not broader research problems. For instance, what is the price elasticity of Uber in different cities; how do messages on Stocktwits influence stock prices; or do people with greater account balances check their banking apps more frequently. These could offer value to companies – but such questions rarely transcend the context (i.e. company) in which they apply. As a field, do we need to be offering tactical advice to companies? They have far superior capabilities to assess their own data – and can do it better. The distinct competence of academic research is to abstract and generalize – so that papers are cited for the broader knowledge product, and yes, even derivative practical interventions (Hirschheim, 2019). Big data has tremendous value in providing insights, albeit tactical ones, that can be abstracted. However,
abdicating knowledge in the form of theory would relegate a field to at best consulting knowledge or at worst a dustbin of empiricism.

The key point here for practice is that knowledge management involves identifying and codifying knowledge that is replete with experience and expertise. Big data offers tremendous potential, but companies that see it as a panacea to replace the insights and intuition, often accumulated through education and experience would be a mistake. So human judgment and domain expertise are needed on both the input and output sides of big data and its analysis – to frame the right questions, access the right data sources, and to interpret the veracity and viability of the results for decision-making. So, companies that can create synergies between their KMS and their big data and analytical systems are more likely to do better. This can occur in many ways, and in both directions. Knowledge codified in a KMS can be subject to enhancement through big data analysis. For example, externalizing knowledge from top salespeople in a KMS on what are their points of emphasis for selling successfully to customers can be enhanced through text analysis of customer feedback or reviews on what customers like. Similarly, strong prescriptive implications emerging from big data analysis can be represented as knowledge in the KMS. A trucking company running optimization algorithms across massive data in its fleet might derive useful prescriptions for fuel conservation that can be codified in a KMS.

For IS researchers, big data and analytics could provide important insights and predictions. But these do not replace theory, but feed into theory, and so the hard work of abstracting from these insights to a general archetypical problem is important to building cumulative knowledge (Rai, 2017). Moreover, big data can offer critical value in triangulating results and in opportunities for conducting digital (field) experiments.

In sum, big data is not knowledge, we should be wary of practices in both business and academia that go overboard on data orientation. This would be like physician treating symptoms directly without a diagnosis (Weick, 1995). Human skills and creativity are essential to complement the big data analytics. So, to truly leverage the power of big data, companies should not simply jump on this technology-driven bandwagon, but carefully develop the multi-level structures and mechanisms so that big data systems can work synergistically with human knowledge. For researchers, driving tactical problems based on accessibility of datasets or tools should not abdicate the need for the hard work of creating knowledge through abstraction. With the exponential growth of digitalization in a global-connected world, big data is only going to increase … and so calibrating our practices now can set the appropriate foundation for knowledge as discriminately valued in the corporate and academic worlds.

Notes on contributor

Varun Grover is the David D. Glass Endowed Chair and Distinguished Professor of IS at the Walton College of Business, University of Arkansas. He has published extensively in the information systems field, with over 400 publications, 250 of which are in major refereed journals. He is Senior Editor for MISQ Executive, Editor of the Journal of the Association for Information Systems Section on Path Breaking Research, and has served as Senior Editor for MIS Quarterly, Journal of the AIS, and Database. Dr. Grover’s current work focuses on the impacts of digitalization on individuals and organizations. He has been extensively involved with PhD students, serving as an advisor to over 40 PhD students. He has been invited to give numerous keynote addresses and talks at various institutions, conferences and forums around the world.

References


Varun Grover

*Walton College of Business, University of Arkansas*

✉️ vgrover@uark.edu